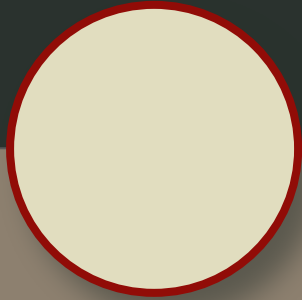


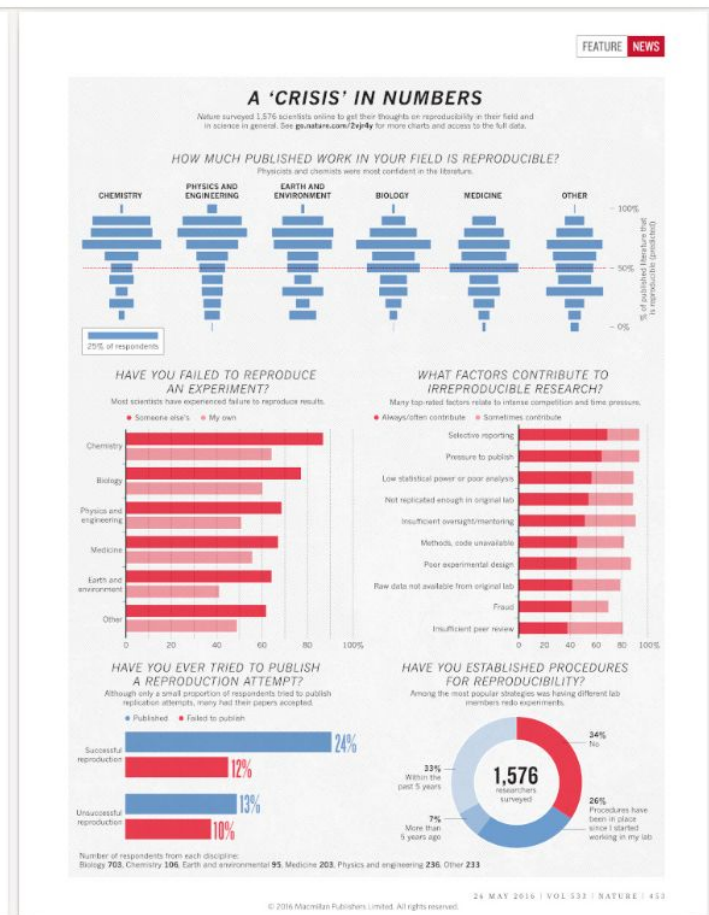
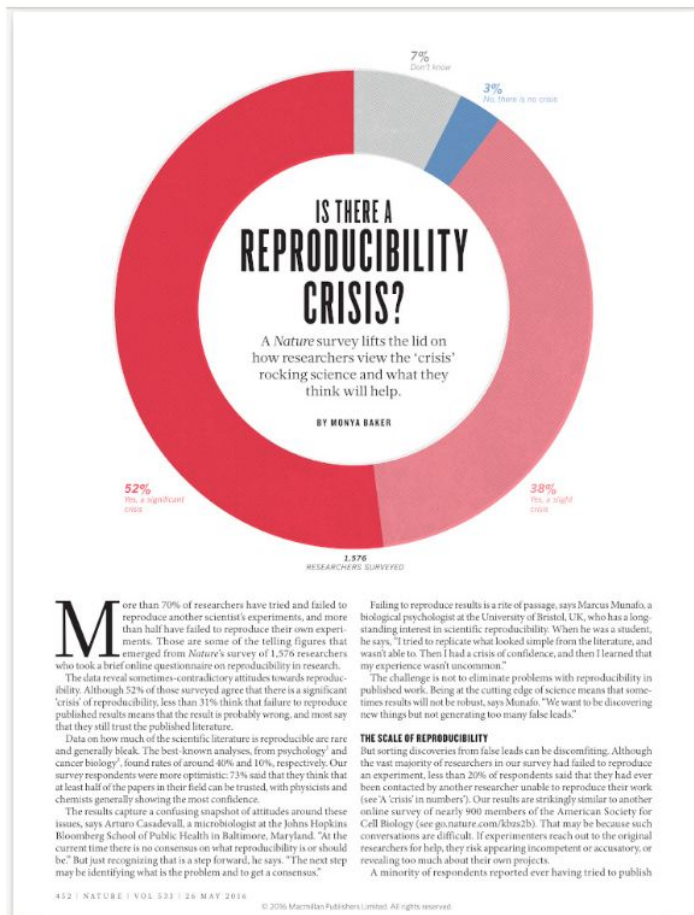
Environments for reproducibility

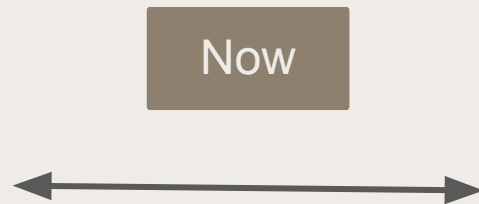


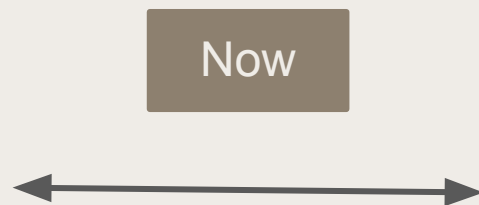
Javier Moldón
IAA-CSIC (Granada)

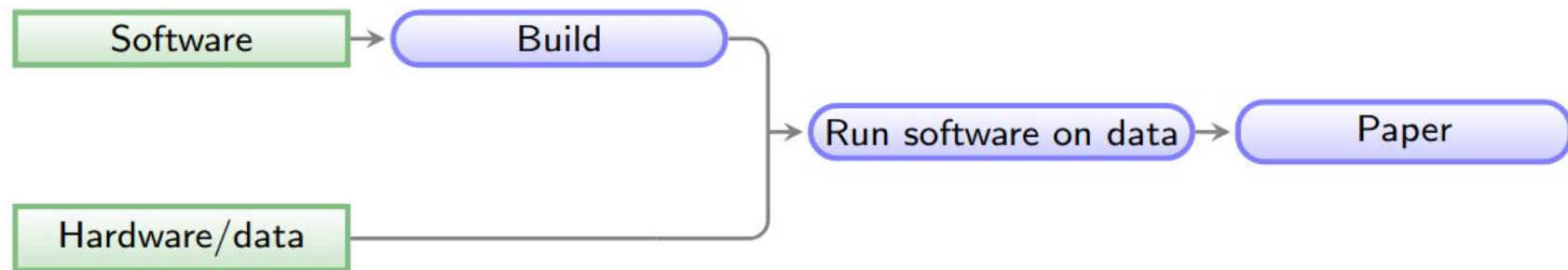
Herramientas para la reproducibilidad del análisis científico
Curso del CSIC
April 20, 2022

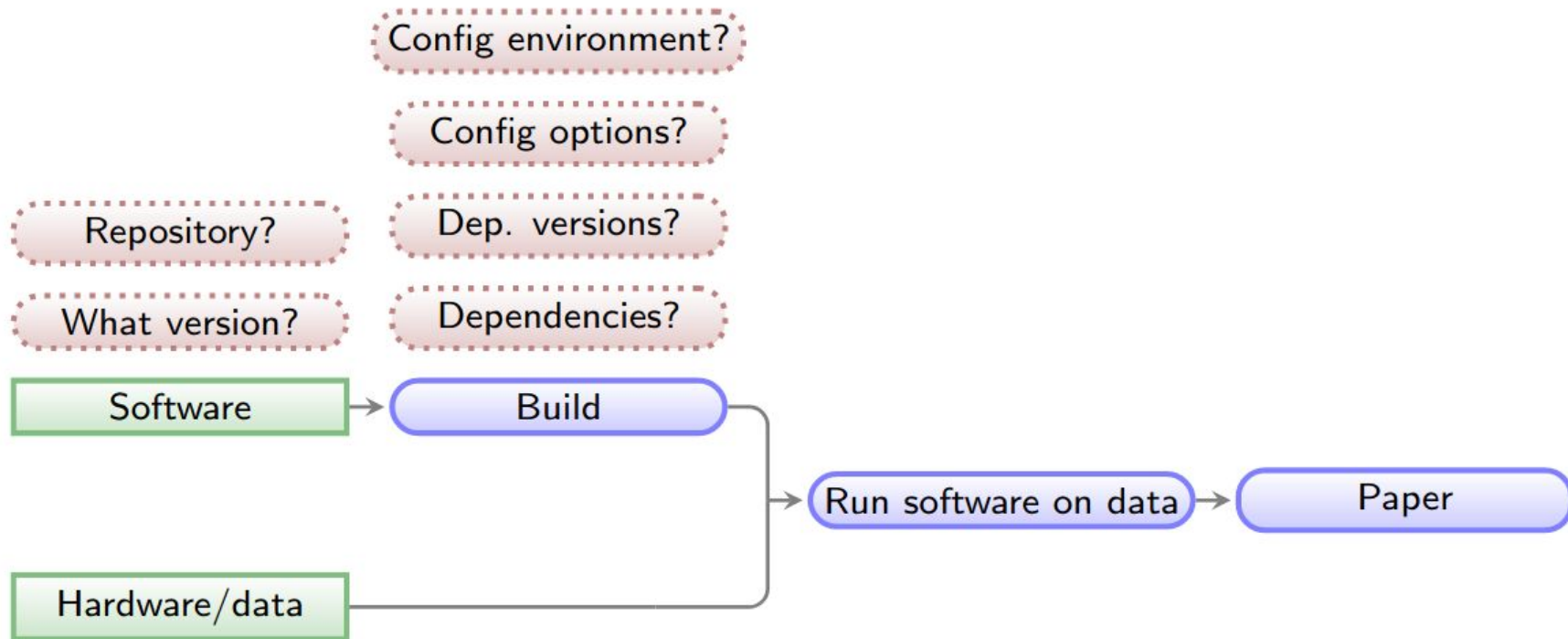
"Reproducibility crisis" in the sciences? (Baker 2016, Nature 533, 452)











Existing solutions:

Virtual machines

Containers (e.g., Docker)

OSs (e.g., Nix, GNU Guix)

Config environment?

Config options?

Repository?

Dep. versions?

What version?

Dependencies?

Software

Build

Runtime options?

What order?

Run software on data

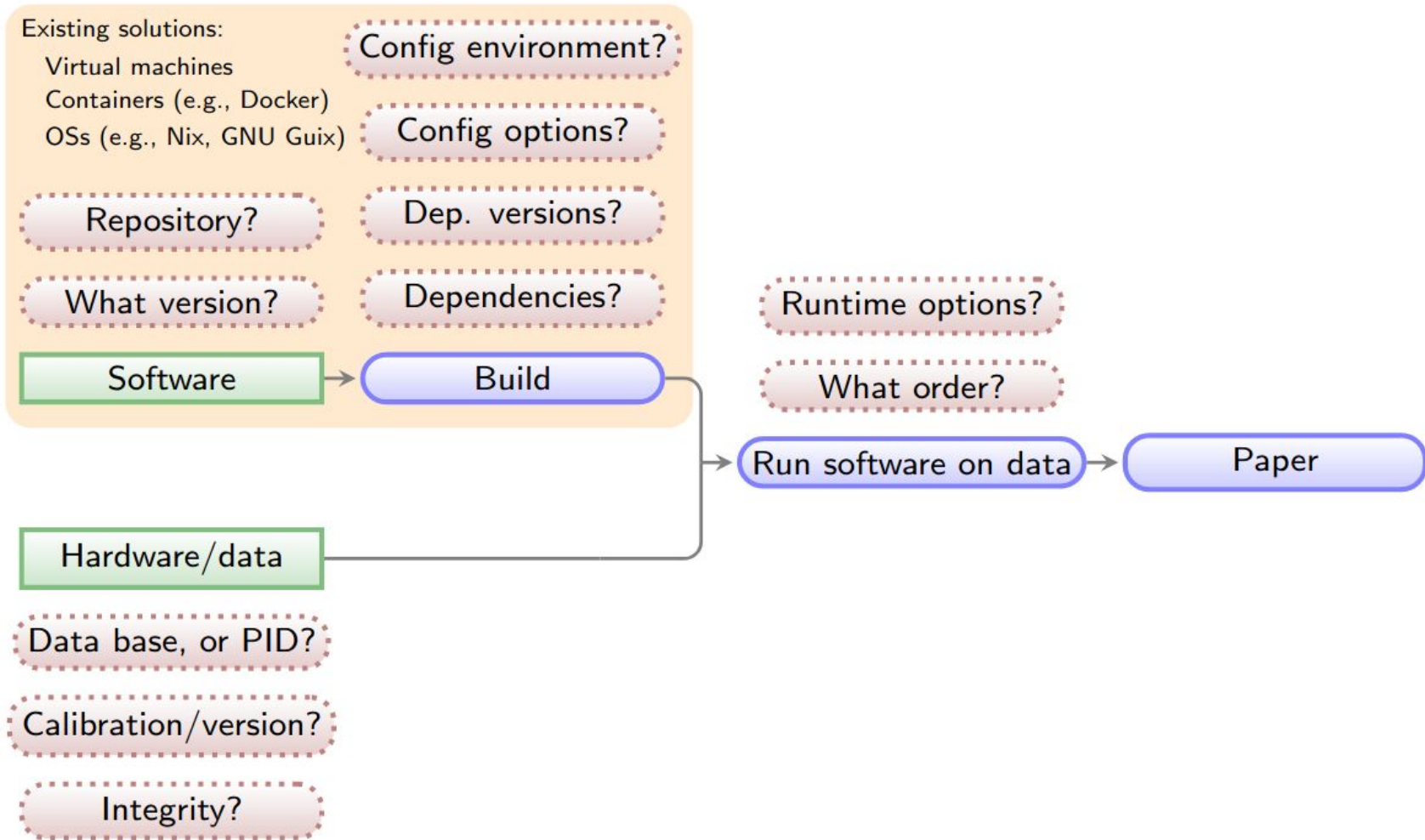
Paper

Hardware/data

Data base, or PID?

Calibration/version?

Integrity?



Software dependencies

The objective is to make your analysis reproducible (by you+everyone)

An analysis may require different packages

- Load and process data
- Visualization
- Statistical analysis
- Advance processing
- ...



Software dependencies

The objective is to
make your
analysis
reproducible (by
you+everyone)

Software dependencies can change over time

An analysis that works today most probably will not work in a few years time. For example `py2` → `py3`.



Tracking all dependencies is very hard!

Some options:

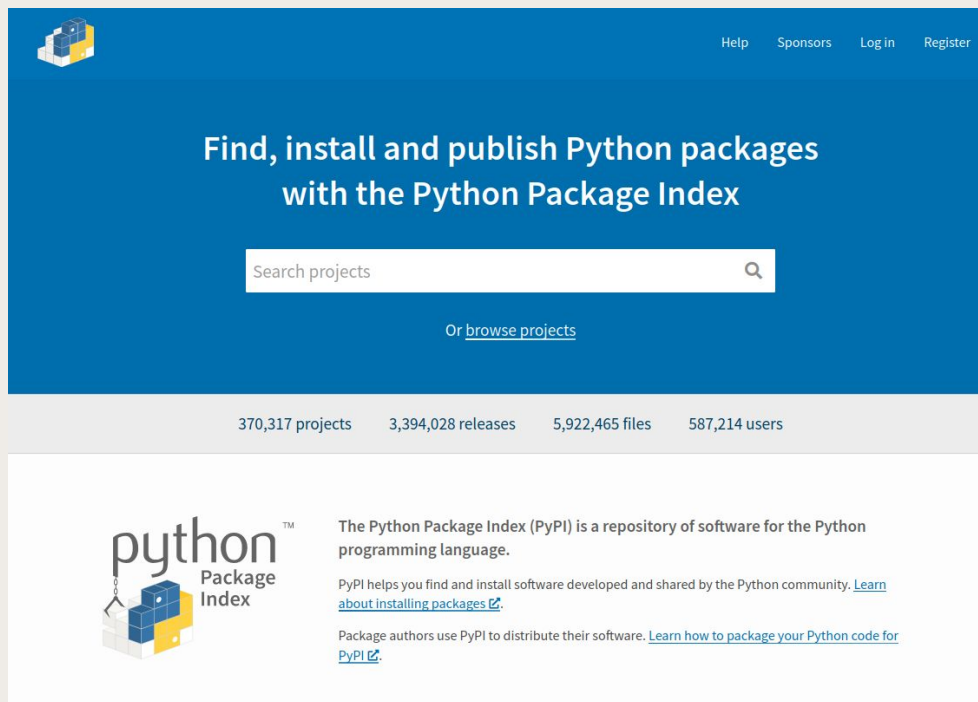
- Write in the README the list of dependencies and versions
- Use a package manager (pip, Packrat, conda, ...)
 - Explicitly fix the dependencies and versions
 - Incorporates a way to install them
- Containers (singularity/docker)
 - Fixed “virtual machine”, like a self-contained computer
 - Almost like a black box. Cannot be modified

pip

pip is the [package installer](#) for Python. You can use pip to install packages from the [Python Package Index](#) and other indexes.

pip

<https://pypi.org/>



Conda / Anaconda / miniconda

Why conda

What is conda?

Conda is a package manager used in scientific computing. It provides scientific libraries and dependencies.

Why conda?

- Manage the software for a project
- Can have different versions for each project
- You create virtual environments, encapsulated and reproducible

Conda

Conda

<https://docs.conda.io/projects/conda/en/latest/index.html>

Package, dependency and environment management for any language---Python, R, Ruby, Lua, Scala, Java, JavaScript, C/ C++, FORTRAN

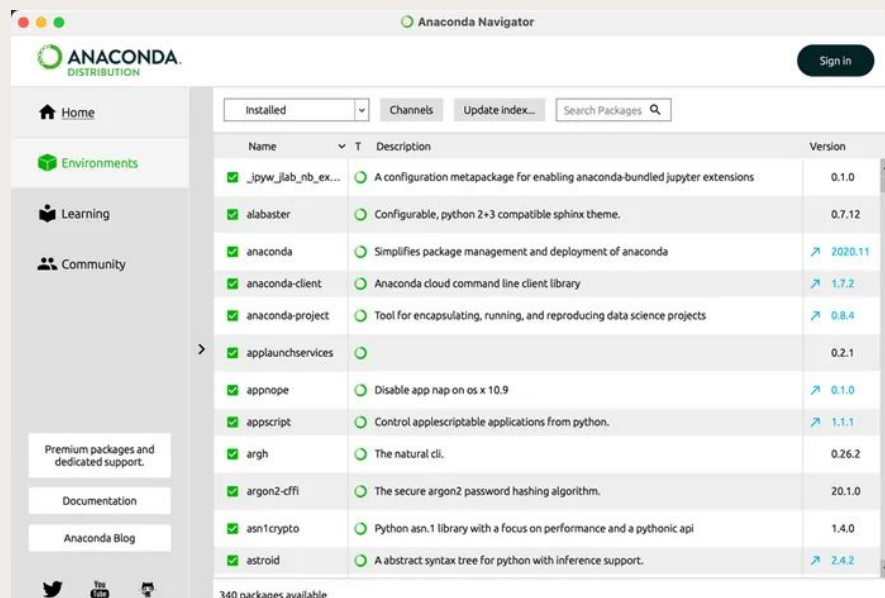
The logo for Conda, featuring a large green 'C' with a white circular pattern inside, followed by the word 'ONDA' in a bold, green, sans-serif font.

Anaconda

A private company providing software and package management solutions

Anaconda distribution

<https://www.anaconda.com/products/distribution>



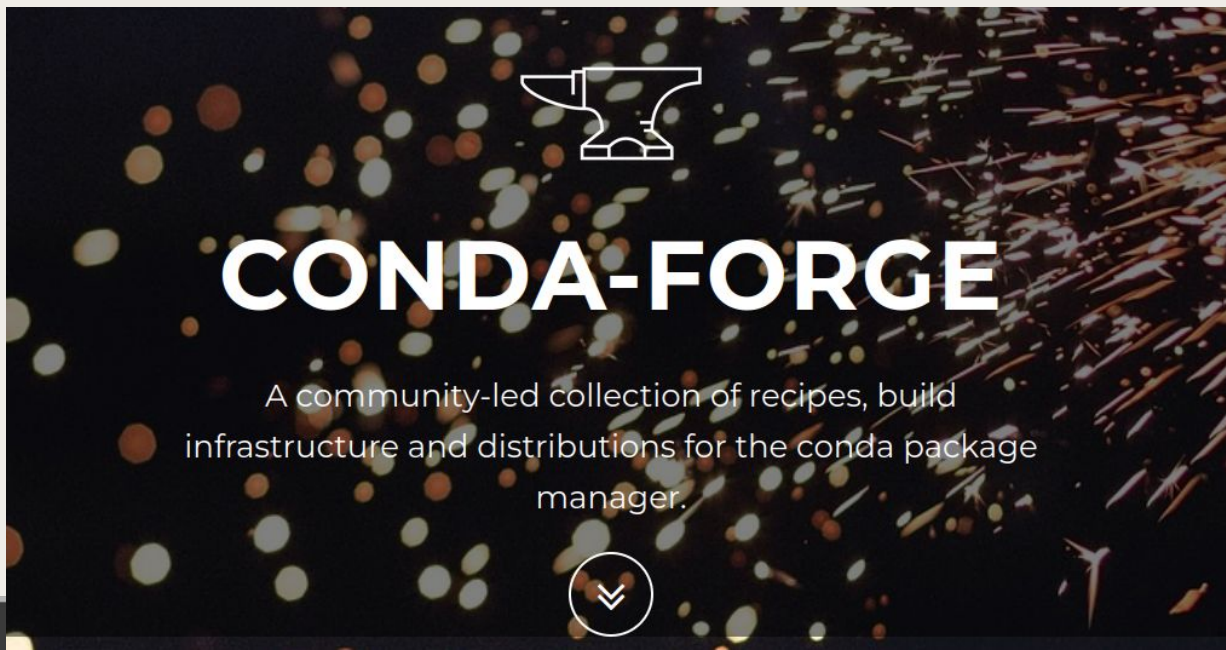
conda-forge

A community-led collection of recipes, build infrastructure and distributions for the conda package manager

Package repository

<https://conda-forge.org/>

<https://github.com/conda-forge>



anaconda.org

Search available packages

<https://anaconda.org/>



Where packages, notebooks, projects and environments are shared.

SEARCH PACKAGES

Efficiency: miniconda mamba

miniconda

A lightweight version of conda. Very easy to install

<https://docs.conda.io/en/latest/miniconda.html>

mamba

A very fast dependency solver. Change conda → mamba

<https://anaconda.org/conda-forge/mamba>

```
conda install -c conda-forge mamba
```

Demo conda